MASTER'S THESIS

Learning analytics als instrument voor vroege detectie van eerstejaars risicostudenten door docenten: een haalbaarheidsstudie.

Chocolaad, Rosanne

Award date:
2021

Open
Universiteit

Learning analytics als instrument voor vroege detectie van eerstejaars risicostudenten door docenten: een haalbaarheidsstudie

Learning analytics as an instrument for lecturers for early detection of first year at-risk students: a feasibility study

Rosanne Chocolaad

Master Onderwijswetenschappen
Open Universiteit

Cursusnaam en cursuscode: OM9906 - Mastherthesis
Naam begeleider: dr. J.P.W. Janssen
Datum: woensdag 30 juni 2021

Inhoudsopgave

Learning analytics als instrument voor vroege detectie van eerstejaars risicostudenten door docenten: een haalbaarheidsstudie

R.E. Chocolaad

Samenvatting

Zowel de strategische agenda van de Vereniging Hogescholen als de agenda van de Hogeschool van Amsterdam is erop gericht om studie-uitval te beperken. Hiertoe zijn op opleidingsniveau reeds verschillende maatregelen getroffen zoals ingangsselectie, bindend studieadvies en zorgvuldige onderwijs- en toetsprogrammering. Daarnaast kan een docent op basis van signalen of feedback die hij ontvangt van een student – bijvoorbeeld tijdens de interactie met studenten in de les of bij langdurige afwezigheid van een student – ook het onderwijsproces bijstellen en studenten bijsturen. Mogelijk kan een docent daarbij ook gebruik maken van door studenten gegenereerde data die real-time en automatisch wordt verzameld in de onderwijssystemen waarmee gewerkt wordt. Er is echter nog weinig onderzoek gedaan naar hoe een docent op basis van door de student gegenereerde data in onderwijssystemen in een zo vroeg mogelijk stadium studenten kan identificeren die mogelijk de onderwijseenheid niet zullen behalen en baat zouden hebben bij een interventie. Learning analytics biedt hier mogelijk een oplossing voor.

Het doel van dit onderzoek is om na te gaan of en op welke wijze data uit de onderwijssystemen van de opleiding Commerciële Economie voor learning analytics doeleinden gebruikt kunnen worden, om op het niveau van onderwijseenheid docenten te informeren over studenten die mogelijk de onderwijseenheid niet zullen behalen. Hiertoe zijn op basis van de in de literatuur gevonden voorspellers en de beschikbare data in de onderzoekscontext de volgende predictoren voor dit onderzoek geoperationaliseerd: het totaal aantal bezochte pagina's in de digitale leeromgeving, de daaraan bestede tijdsduur en de aanwezigheid in de les voor de gehele onderwijsperiode. Ook werd er specifiek gekeken naar de data van de eerste drie lesweken, om vroegtijdig risicostudenten te kunnen signaleren. Daarvoor werden de vroegtijdige voorspellers geoperationaliseerd: het totaal aantal bezochte pagina's van de betreffende eerste drie lesweken, de daaraan bestede tijdsduur en de aanwezigheid in de les.

Via een retrospectief onderzoek is met secundaire data van 349 studenten door middel van een logistische regressie analyse de relatieve voorspellende waarde van de verschillende variabelen vastgesteld. Over de gehele onderwijsperiode blijkt de totale aanwezigheid en het totaal aantal bezochte pagina's het wel of niet behalen van de module te voorspellen. Echter wordt van de groep studenten die het vak niet behaalt (n = 121) slechts in 27.3% van de gevallen een correcte voorspelling gemaakt. Waar het gaat om vroegtijdige signalering van de risicostudenten, blijkt dat alleen de aanwezigheid in de eerste drie lesweken een significante voorspeller is, maar ook hier wordt voor slechts 27.3% van de studenten die het vak niet behalen een correcte voorspelling gemaakt.

Alhoewel met dit model en deze voorspellers de opleiding dus wel een deel van de risicostudenten via learning analytics in kaart zou kunnen brengen en een interventie zou kunnen plegen, is het wenselijk om het model eerst te optimaliseren om betere voorspellingen te doen. Het huidige model biedt onvoldoende handvatten om op het niveau van onderwijseenheid docenten vroegtijdig te informeren over studenten die het risico lopen de module niet met succes af te ronden, maar het onderzoek biedt interessante perspectieven en bevindingen om verder binnen de opleiding en faculteit de dialoog met elkaar aan te gaan over learning analytics en de mogelijkheden die learning analytics biedt, verder te verkennen.

Trefwoorden: learning analytics, aanwezigheid, digitale leeromgeving, risicostudenten, vroege detectie, voorspellen

Learning analytics as an instrument for lecturers for early detection of first year at-risk students: a feasibility study

R.E. Chocolaad

Summary

Both the strategic agenda of the Vereniging Hogescholen and the Amsterdam University of Applied Sciences aim at limiting dropout rates. To this end, various measures have already been taken at program level, such as entrance selection, binding study advice and conscientious teaching and test programming. In addition, a lecturer can also adjust the teaching process and guide students based on signs and feedback from the students for example, during interaction with students in class or in the event of a student's long-term absence. This can possibly be supplemented with real-time and automatically collected data generated by students in the educational systems with which they work. However, little research has been done into how a lecturer, based on data generated by students in educational systems, can identify students who may not pass the course at an early stage and might benefit from an intervention. Learning analytics may offer a solution for this.

The aim of this research is to determine whether and how data from the educational systems of the CE study program can be used for learning analytics purposes to inform lecturers about students who may not pass a course. To this end, based on the predictors found in the literature and the available data in the research context the following predictors for this research were operationalized: the total number of learning management system pages visited, the time spent on them and attendance in class for the entire teaching period. The data of the first three weeks of classes were also specifically looked at, in order to be able to identify at risk students at an early stage and the following early predictors were operationalized: the total number of pages visited concerning the first three teaching weeks, the time spent on them and the attendance in the lesson.

Through a retrospective study, secondary data from 349 students was used to determine the relative predictive value of the various variables by means of a logistic regression analysis. Over the entire teaching period, only the total attendance and the total number of pages visited appears to predict whether or not a student passed the module. However, of the group of students who fail the course (n = 121), a correct prediction is made only in 27.3% of the cases. When it comes to early identification of these at risk students, it appears that only attendance in the first three weeks of lessons is a significant predictor, but here too, a correct prediction is made for only 27.3% of the students who do not pass the course.

Although with this model and these predictors a number of at risk students could be predicted and possibly intervened with, it is desirable that the model is first further optimized in order to make better predictions. The current model offers insufficient input to inform lecturers at an early stage about students who run the risk of not successfully completing the course, but the research presents

interesting perspectives and findings to further discuss and explore learning analytics and the possibilities it has to offer within the study program and faculty.

Keywords: learning analytics, attendance, learning management system, at risk student, drop out, early detection, prediction

1. Inleiding

In het hoger onderwijs in Nederland wordt studie-uitval breed als probleem ervaren. Zowel de strategische agenda van de Vereniging Hogescholen als de agenda van de Hogeschool van Amsterdam (HvA) is erop gericht om studie-uitval te beperken (Hogeschool van Amsterdam, 2018; Vereniging Hogescholen, 2019). Studenten die kiezen voor de HvA opleiding Commerciële Economie (CE) maken in hun eerste studiejaar kennis met belangrijke begrippen uit de Marketing en Sales wereld. Van de ongeveer 800 studenten die jaarlijks in september starten met de opleiding, stromen aan het einde van het studiejaar ongeveer 400 à 500 studenten door naar het tweede jaar. Uit het management informatie systeem van de instelling blijkt dat ongeveer 40% van de studenten stopt met de opleiding als gevolg van een negatief bindend studieadvies (BSA). Dit heeft niet alleen belangrijke consequenties voor de studenten, maar ook voor de onderwijsinstelling en de maatschappij in het algemeen. Een student investeert namelijk tijd, energie en geld – vaak in de vorm van een lening – in de studie. Door uitval zijn deze investeringen van de student teniet gedaan. Hetzelfde geldt voor de onderwijsinstelling die ook tijd en middelen investeert in de student.

Er is al het nodige onderzoek gedaan naar studie-uitval (Delnoij, Dirkx, Janssen, & Martens, 2020). In hun reviewstudie komen Delnoij et al. (2020) tot een raamwerk waarin de oorzaken geclassificeerd worden. Daarbij maken zij onderscheid tussen demografische factoren zoals de sociaaleconomische status van de student, affectieve en cognitieve factoren zoals kennis en ervaring en situationele factoren, zoals bijvoorbeeld mantelzorgverplichtingen van de student. Ook kunnen institutionele factoren, gekoppeld aan het beleid en de procedures van de onderwijsinstelling, een rol spelen. Op het niveau van de onderwijsinstelling en opleiding kunnen verschillende maatregelen worden genomen om uitval te voorkomen en te verminderen zoals ingangsselectie en zorgvuldige onderwijs- en toetsprogrammering (Van Berkel, Jansen, & Bax, 2012). Daarnaast kan een docent op basis van signalen die hij of zij ontvangt van een student – bijvoorbeeld tijdens de interactie met studenten in de les of bij langere afwezigheid van een student – ook het onderwijsproces bijstellen en studenten bijsturen. Mogelijk kan een docent daarbij ook gebruik maken van door studenten gegenereerde data die real-time en automatisch wordt verzameld in de onderwijssystemen waarmee gewerkt wordt. Er is echter nog weinig onderzoek gedaan naar hoe een docent op basis van deze data in een zo vroeg mogelijk stadium studenten kan identificeren die mogelijk de onderwijseenheid of module niet zullen behalen en baat zouden hebben bij een interventie (Gray & Perkins, 2019). Het niet succesvol afronden van een onderwijseenheid kan namelijk bijdragen aan een negatief studieadvies en studie-uitval in het eerste jaar. Door docenten op het niveau van een onderwijseenheid een instrument aan te bieden waarmee zij studenten kunnen signaleren die mogelijk de onderwijseenheid niet zullen behalen, kunnen zij vroegtijdig het gesprek met studenten aangaan en ze beter begeleiden en/of de leer- en instructiestrategieën aanpassen.

Learning analytics biedt hier mogelijk een oplossing voor. Learning analytics is het verzamelen, analyseren en rapporteren van data uit de leeromgeving ten behoeve van het verbeteren van het leerproces van studenten (SURF, 2019). Door learning analytics kan informatie die voorheen onopgemerkt en onzichtbaar bleef worden ontsloten. Mogelijke patronen die studie-uitval voorspellen kunnen worden afgeleid op basis waarvan docenten actie kunnen ondernemen (Bienkowski, Feng & Means, 2012; Chatti, Dyckhoff, Schroeder, & Thüs, 2012). Steeds meer hoger onderwijsinstellingen gebruiken learning analytics om studiesucces te verbeteren (Scheffel, Drachsler, Stoyanov, & Specht, 2014). Norris en Baer (2013) stellen zelfs dat het implementeren van learning analytics door een onderwijsinstelling essentieel is voor studiesucces. Hier moet binnen de specifieke onderwijscontext wel onderzoek naar worden verricht; learning analytics is namelijk geen one-size-fits-all oplossing. Doordat eerder onderzoek is uitgevoerd in verschillende contexten is het niet eenvoudig om de uitkomsten met elkaar te vergelijken of algemene conclusies te trekken (Conijn, Snijders, Kleingeld, & Matzat, 2017). Bij de opleiding CE wordt momenteel data verzameld in verschillende onderwijssystemen, waaronder een digitale leeromgeving en een aanwezigheidsregistratiesysteem. Deze data herbergen potentieel een schat aan informatie die tot op heden niet is onderzocht, laat staan gebruikt. Het doel van dit onderzoek is om na te gaan of en op welke wijze de data uit deze systemen voor learning analytics doeleinden gebruikt kunnen worden, om op het niveau van een onderwijseenheid docenten te informeren over studenten die mogelijk de onderwijseenheid niet zullen behalen.

1.1 Theoretische kader

In deze paragraaf worden de theoretische achtergrond en eerdere onderzoeken gepresenteerd die verband houden met het doel van dit onderzoek. Daarbij wordt, indien van toepassing, meteen ook de afbakening naar de eigen onderzoekspraktijk gemaakt.

1.1.1 Studieuitval in het eerste jaar

Jaarlijks starten in september ongeveer 800 studenten aan een nieuw avontuur: een studie Commerciële Economie (CE) aan de HvA. Uit de cijfers van de instelling blijkt dat velen al vroeg in het avontuur stranden: ongeveer 40% van de studenten stopt met de opleiding omdat zij niet het minimum aantal vereiste studiepunten voor het eerste jaar hebben behaald. In het eerste studiejaar kan een HvA-student 60 studiepunten behalen. De HvA hanteert een BSA-norm van 50 studiepunten. Een student die minder dan 50 studiepunten behaalt, ontvangt een negatief studieadvies en mag de opleiding niet vervolgen. Zoals gesteld in de Wet op het hoger onderwijs en wetenschappelijk onderzoek is het studieadvies bedoeld om in een vroeg stadium vast te stellen of een student geschikt is voor een opleiding. De achterliggende gedachte daarbij is dat een student zich eventueel kan oriënteren op een andere studie (Van den Ende, De Vreede, Zandvliet, & de Vleeschouwer, 2019). Bij de opleiding CE worden de studiepunten per module of onderwijseenheid toegekend. Het aantal

studiepunten kan variëren tussen 4, 5 en 7 studiepunten. Een module bestaat uit meerdere deeltoetsen (tentamens, opdrachten, verslagen, etc.) en voor sommige onderdelen geldt een ondergrens in de vorm van een minimumcijfer. De studiepunten voor de module worden toegekend als het gewogen eindcijfer onafgerond ten minste een 5,5 is.

Het doel van de BSA-norm is dus om studenten eruit te filteren waarvoor geldt dat de gekozen opleiding qua niveau en/of inhoud geen goede 'match' vormt, maar het is de vraag of dit ook echt het geval is voor alle BSA-uitvallers, in dit geval dus 40% van de studenten. Zeker gezien het feit dat zij aan de vooropleidingseisen hebben voldaan. Het zou hier bijvoorbeeld ook om studenten kunnen gaan die moeten wennen aan de overstap havo/mbo of buitenlandse vooropleiding naar het hbo, studenten met (te veel) nevenactiviteiten of andere persoonlijke omstandigheden. In het rapport dat Van den Ende et al. (2019) in opdracht van het Ministerie van Onderwijs, Cultuur en Wetenschap hebben opgesteld is geconcludeerd dat vroegtijdige signalering en daarop afgestemde begeleiding belangrijk is om onnodige uitval van geschikte studenten te voorkomen. Als een docent reeds in een vroeg stadium kan signaleren dat een student de module mogelijk niet zal behalen, kan hij of zij tijdig het gesprek aangaan met de student en waar nodig interventies plegen. Als de student de module reeds onvoldoende heeft afgerond, is het hier soms al te laat voor. Voor het vroegtijdig signaleren van deze risicostudenten biedt learning analytics mogelijk een oplossing en daarbij zijn data uit verschillende onderwijssystemen nodig.

1.1.2 Onderwijsdata en learning analytics

In de huidige situatie wordt informatie over studenten, hun studie-ervaringen en de kwaliteit van het onderwijs nog veelal op geaggregeerd niveau afgeleid uit enquêtes uit onderwijsevaluaties, de studieresultaten van de studenten, slagingspercentages van toetsen en toetsresultaten. Deze data wordt achteraf – na afloop van de onderwijsperiode of het semester –geanalyseerd om na te gaan hoe en waar het onderwijs voor de volgende periode verbeterd kan worden en/of effectiever kan worden ingericht (Coates, 2005). Waar het om de leerprestaties van studenten gaat, wordt vooral gekeken naar summatieve toetsmomenten waarmee de onderwijseenheid in de meeste gevallen wordt afgerond. Voor studenten die de module niet met een voldoende afronden, is het dan vaak al te laat om bij te sturen of in te grijpen (Gray & Perkins, 2019). De afgelopen decennia echter, is het onderwijs steeds meer gedigitaliseerd en wordt er gebruik gemaakt van verschillende nieuwe technologische toepassingen en onderwijssystemen zoals bijvoorbeeld de digitale leeromgeving (DLO). Alle interacties van studenten met deze onderwijssystemen worden automatisch geregistreerd en dat maakt dat er anno 2021 continu en real-time grote hoeveelheden data beschikbaar zijn die mogelijk informatie bevatten over het leerproces van de studenten en de mate waarin zij betrokken zijn bij het onderwijs van de module (Lockyer, Heathcote, & Dawson, 2013). Deze data kan als input dienen voor learning analytics doeleinden.

Greller en Drachsler (2012) hebben een raamwerk opgesteld waarin zes verschillende dimensies zijn gedefinieerd waar rekening mee moet worden gehouden bij het ontwerpen van een betekenisvol learning analytics proces in een onderwijsinstelling, namelijk de doelstellingen, belanghebbenden, data, instrumenten en de interne en externe beperkingen. De dimensie doelstellingen heeft betrekking op het doel waarmee learning analytics wordt ingezet. Drachsler en Greller (2012) maken hierbij onderscheid tussen reflecteren en voorspellen. Learning analytics voor reflectie-doeleinden geeft onder andere feedback aan de docent. Door kritisch te kijken naar de beschikbare data kan de docent bijvoorbeeld tot nieuwe inzichten komen over de inrichting en inhoud van de onderwijseenheid, maar ook zien hoe individuele studenten of groepen studenten ervoor staan. Bij learning analytics met als doel voorspellen gaat het vooral om het zo vroeg mogelijk identificeren van studenten, die mogelijk de module niet zullen behalen of afronden en tijdig een interventie te kunnen plegen (Drachsler & Greller, 2012; Van den Bogaard & De Vries, 2017). De dimensie belanghebbenden omvat de data clients en data subjects: De data client ontvangt informatie op basis van data die zijn verzameld door data subjects en kan daar actie op ondernemen. Dit kan op verschillende niveaus: Op nationaal en onderwijsinstellingsniveau kan data gebruikt worden om verantwoording af te leggen over de studievoortgang van cohorten, de studierendementen en voor bijvoorbeeld accreditatiedoeleinden. Op het niveau van de opleiding kan met behulp van learning analytics daarnaast ook het gehele curriculum geanalyseerd worden en op het niveau van een onderwijseenheid of module kan specifiek worden gekeken naar het leerproces van een student (Dietz-Uhler & Hurn, 2013). Binnen dit onderzoek ligt de nadruk op learning analytics op het niveau van een onderwijseenheid en het leerproces van individuele studenten. De dimensie data heeft betrekking op de beschikbare datasets in een onderwijsinstelling. Deze datasets zitten vaak in verschillende onderwijssystemen en met learning analytics wordt de mogelijkheid gecreëerd om deze data te combineren, te analyseren en tot nieuwe inzichten te komen. Echter is een groot deel van deze data beveiligd, moeilijk of niet toegankelijk en niet als zogenoemde open data beschikbaar voor deze doeleinden. Het werken met geanonimiseerde data biedt hiervoor een mogelijke uitkomst. De wijze waarop de data wordt geanalyseerd en beschikbaar wordt gesteld aan de stakeholders rekenen Greller en Drachsler (2012) tot de dimensie instrumenten. Daarbij kan data verkend worden aan de hand van de educational data mining (EDM) techniek of klassieke statistische analyses. Het grootste verschil tussen deze twee benadering heeft vooral te maken met de wijze waarop de data wordt geanalyseerd. Bij EDM wordt niet vraag-gestuurd in grote hoeveelheden data geautomatiseerd gezocht naar patronen en mogelijke verbanden, terwijl er bij de klassieke statistische learning analytics techniek gericht vanuit een vraag of hypothese wordt gewerkt om met behulp van statistische analyses tot een oordeel te komen (Akhtar, Warburton, & Xu, 2017; Bienkowski et al., 2012; Siemens & Baker, 2012). Voor dit onderzoek heeft de klassieke statistische learning analytics techniek de voorkeur omdat er op basis van een vraag data uit verschillende digitale onderwijssystemen wordt verzameld en geanalyseerd. Door erover te rapporteren kunnen nieuwe inzichten verkregen worden over studenten, hun leerproces

en de omgevingen waarin het leren plaatsvindt (Siemens, 2013; Siemens & Baker, 2012). Tot slot noemen Greller en Drachsler (2012) de dimensies: externe en interne beperkingen. De externe beperkingen hebben onder andere betrekking op de ethische en privacy vraagstukken die komen kijken bij het verzamelen en gebruiken van data voor learning analytics-doeleinden in het onderwijs en de interne beperkingen hebben betrekking op de competentie van de data clients: Hoe competent zijn de data clients in het interpreteren van en kritisch nadenken over de gevisualiseerde data?

Learning analytics is niet gestoeld op specifieke onderwijskundige strategieën of epistemologische theorieën. Greller en Drachlser (2012) gaan er in hun raamwerk vanuit dat onderwijskundige strategieën geëvalueerd kunnen worden aan de hand van de beschikbare learning analytics data en daardoor het effect van die strategieën op leren en instructie vastgesteld kan worden. Het vormt echter wel een uitdaging om indicatoren te formuleren in de beschikbare datasets die het meest relevant zijn voor het evalueren van een leer- en instructieproces dat gestoeld is op bepaalde onderwijskundige principes (Greller & Drachsler, 2012; Lockyer et al., 2013). Er wordt nog veel onderzoek gedaan naar learning analytics toepassingen in het onderwijs. Dit gebeurt in verschillende contexten – soms met gelijke, maar soms ook met verschillende variabelen, waardoor het niet eenvoudig is om de uitkomsten met elkaar te vergelijken of algemene conclusies te kunnen trekken (Conijn et al., 2017). Learning analytics beslaat dus een breed gebied, maar in het kader van dit onderzoek worden in de volgende paragraaf enkel resultaten uit voorgaand onderzoek gepresenteerd die verband houden met voorspellers van studieresultaten en studie-uitval. Hierbij ligt de nadruk op voorspellers die tijdens het onderwijsproces worden gegenereerd in de voor dit onderzoek toegankelijke digitale onderwijssystemen.

1.1.3 Voorspellers van studieresultaten en studie-uitval

Studie-uitval is een belangrijk, wijd bestudeerd en actueel onderzoeksthema binnen het hoger onderwijs (McCoy & Bryne, 2017; Paige, Wall, Marren, Dubenion, & Rockwell, 2017). Sinds de jaren 70 zijn er door verschillende onderzoekers diverse conceptuele modellen samengesteld met variabelen die van invloed zijn op studie-uitval. Het model van Tinto (1975) vormt daarbij een belangrijke leidraad. In het model zijn niet alleen variabelen opgenomen die betrekking hebben op de eigenschappen, achtergrond en vooropleiding van de student, maar ook de academische prestaties en de sociale integratie en binding van de student met de opleiding of onderwijsinstelling. Op basis van (revisies van) het model van Tinto is er, aan de hand van deze factoren, reeds veel onderzoek gedaan naar studie-uitval. Anno 2021 zijn er, als gevolg van de digitalisering in het onderwijs, ook continu en real-time grote hoeveelheden data beschikbaar in verschillende onderwijssystemen. De opkomst van learning analytics maakt het mogelijk om, op het gebied van het voorspellen van studie-uitval, ook deze data als input mee te nemen in onderzoek en een rijker model te creëren (Archer & Prinsloo, 2020).

Zo hebben Arnold en Pistilli (2012) *Course Signals* ontwikkeld: een systeem dat door middel van een voorspellend algoritme studenten kan signaleren die dreigen achter te lopen of de module niet te behalen. De student kan dit zien in de vorm van een verkeerslicht, waarbij rood aangeeft dat de kans groot is dat het vak niet behaald wordt. De docent heeft deze informatie ook tot zijn beschikking en kan een interventie plegen. Arnold en Pistilli (2012) hebben als input voor *Course Signals* een algoritme ontwikkeld dat onder andere gebruik maakt van de studieresultaten en de interactie van de student met de DLO, informatie over de vooropleiding en daarin behaalde gemiddelde cijfers, en studentkarakteristieken zoals woonplaats, leeftijd etc. Kortom, een uitgebreid model dat gebruik maakt van veel verschillende gegevens om de voorspellingen te kunnen doen. Er zijn binnen een onderwijsinstelling immers grote hoeveelheden gegevens beschikbaar die gebruikt kunnen worden om de studieresultaten van studenten te voorspellen, maar indachtig het advies van Dietz-Uhler en Hurn (2013) richt het huidige onderzoek zich op eenvoudige vormen van data waar de faculteit over beschikt – en verder afgebakend naar data waar de docent en student invloed op kunnen uitoefenen (Yu & Jo, 2014). Immers, data met betrekking tot de vooropleiding, eerder behaalde studieresultaten en studentkarakteristieken zijn vaststaande gegevens. Bij de opleiding CE wordt data verzameld in verschillende onderwijssystemen: de DLO, een aanwezigheidsregistratie tool en het Studenten Informatie Systeem (SIS). Deze onderwijssystemen en de daarin beschikbare data stellen de kaders voor het huidige onderzoek.

Er zijn reeds verschillende onderzoeken uitgevoerd waarbij is nagegaan in hoeverre DLO-gebruiksdata de studieresultaten van studenten kan voorspellen. In de diverse studies zijn er niet alleen verschillende DLO's gebruikt, maar ook de predictorvariabelen verschillen (Conijn et al., 2017). Reeds in 1997 deden Rafaeli en Ravid (1997) onderzoek naar de relatie tussen studieresultaten en DLO-activiteit van studenten. Zij vonden een positieve relatie tussen de behaalde studieresultaten en het aantal gelezen DLO-pagina's en de resultaten van online oefenvragen. Ook verklaarde de DLO-activiteit en resultaten van online oefenvragen 22% van de variantie in het eindcijfer. Morris, Finnegan en Wu (2005) gebruikten voor hun onderzoek meerdere variabelen en uit hun onderzoek bleek dat het aantal discussieberichten dat een student bekeek, de tijd die hij hieraan besteed had en het aantal bekeken DLO-pagina's significante voorspellers waren van het eindcijfer. Davies en Graff (2005) echter vonden geen relatie tussen de door de student bekeken en geplaatste discussieberichten en het eindcijfer.

Macfadyen en Dawson (2010) vonden een positieve correlatie tussen het aantal links en bestanden dat een student in de DLO had bekeken en het behaalde eindcijfer. Deze variabelen bleken in hun model echter geen significante voorspellers van het eindresultaat en komt overeen met de bevindingen van You (2016). Ryabov (2012) en Firat (2016) vonden daarentegen wel dat de tijdsduur die studenten online in de DLO doorbrachten een significante voorspeller was van het eindcijfer. Yu en Jo (2014) deden ook onderzoek naar DLO-variabelen als voorspellers van prestaties van studenten. Het aantal logins in de DLO en de interactie met docenten bleken geen significante voorspellers van

het behaalde eindcijfer, maar de totaal bestede tijdsduur in de DLO, de regelmaat waarmee de DLO bezocht werd, het aantal leermaterialen dat gedownload was en de interactie met medestudenten wel. Als het gaat om vroegtijdige signalering van risicostudenten vonden Milne, Jeffrey, Suddaby en Higgens (2012) een verband tussen het gebruik van de DLO – en dan met name de bezochte DLO pagina's – in de eerst lesweek en het behalen van het vak: van studenten die het vak niet succesvol hadden afgerond was het DLO gebruik in de eerste lesweek significant lager dan van studenten die het vak wel succesvol hadden afgerond.

Tabel 1 vat de resultaten uit eerdere studies samen. Zoals uit de tabel blijkt, zijn de resultaten uit eerdere studies verschillend en bij gelijke predictorvariabelen soms zelfs ook tegenstrijdig. Dit komt mogelijk door de verschillende contexten en omstandigheden waaronder de data is verzameld en het onderzoek is uitgevoerd. Het vergelijken van deze resultaten is daardoor ook problematisch; deze zijn namelijk nauw verbonden met de inhoud van het onderwijs en daaraan gerelateerde kenmerken zoals de mate waarin en de manier waarop online inhoud wordt aangeboden en bijvoorbeeld discussie expliciet onderdeel uitmaakt van de leeractiviteiten. Als gevolg hiervan kunnen er geen algemene conclusies worden getrokken over de beste en meest geschikte DLO-voorspellers van studieresultaten (Conijn et al., 2017). Onder andere Larrabee Sønderlund, Hughes en Smith (2019), Conijn et al. (2017) en Yu en Jo (2014) geven aan dat verder onderzoek in eigen context nodig is om een degelijke wetenschappelijke basis te vormen. Een andere kanttekening bij deze onderzoeken is dat sommige voorspellers met de nodige voorzichtigheid geïnterpreteerd moeten worden. Het is bijvoorbeeld niet met zekerheid te zeggen of een student gedurende een in de DLO geregistreerde tijdsduur ook daadwerkelijk bezig is geweest met het leermateriaal. Een student kan een bepaalde DLO-pagina bezoeken, maar is tussendoor misschien ook regelmatig afgeleid door social media of ergens anders mee bezig.

Wat betreft de aanwezigheid van studenten in de les als voorspeller van studieresultaten is er ook al het nodige onderzoek verricht. Gray en Perkins (2019) hebben op basis van aanwezigheidsdata van studenten een model ontwikkeld dat reeds in een heel vroeg stadium, in de derde lesweek, met 97% betrouwbaarheid de studenten kan signaleren die mogelijk een vak niet zullen behalen. In het hoger onderwijs is de fysieke aanwezigheid van studenten in de les over het algemeen laag: Vooral nu met alle recente technologische ontwikkelingen, waarbij leren niet meer beperkt is tot het leslokaal, maar ook online plaatsvindt (Brennan, Sharma, & Munguia, 2019). Alhoewel fysieke aanwezigheid van de student in de les niet direct iets zegt over het daadwerkelijke leren, biedt het de docent wel de mogelijkheid om de student op de volgens hem of haar meest geschikte, didactisch verantwoorde manier kennis te laten maken met de leerstof (Brennan, et al., 2019). Daarbij is uit verschillende onderzoeken gebleken dat de fysieke aanwezigheid in de les wel degelijk van invloed is op de studieresultaten – en daarmee op het behalen van een onderwijseenheid. Gurung, Weidert en Jeske (2010) hebben onderzocht welk studiegedrag geassocieerd is met hogere eindcijfers. Daaruit bleek

Tabel 1

Onderzoeksresultaten: Potentiële voorspellers

Variabele	Resultaten & referenties
Aantal DLO-pagina's bezocht	- Positieve correlatie tussen het aantal bezochte pagina's en het eindcijfer. Het aantal bezochte pagina's en de cijfers voor online oefenvragen bepaalde 22% van de variantie in het eindcijfer (Rafaeli & Ravid, 1997). - Positieve correlatie met het eindcijfer, maar geen voorspellende waarde. (Macfadyen & Dawson, 2010) - Een significant verschil tussen het DLO gebruik in de eerste lesweek van studenten die het vak niet succesvol hadden afgerond ten opzichte van studenten die het vak wel succesvol hadden afgerond (Milne et al., 2012) - Aantal bezochte pagina's was één van de statistisch significante voorspellers van het eindcijfer. Samen met het aantal bekeken discussieberichten en de hieraan bestede tijd verklaarde dat 31% in het eindcijfer. (Morris et al., 2015) - Een significante correlatie tussen het aantal bekeken pagina's/bronnen en het eindcijfer. Het bleek één van de meest stabiele voorspellers van het behaalde cijfer. (Conijn et al.,2017)
Tijdsduur online in DLO	- Significante, maar lage correlatie met het eindcijfer en geen voorspeller van het eindcijfer. (Macfadyen & Dawson, 2010) - Significante, positieve relatie tussen de totale tijd die een student online in de DLO doorbracht en het eindcijfer. (Rybov, 2012) - Significante correlatie tussen de tijdsduur online in de DLO, online interactie met peers en het eindcijfer. Het model waarin deze variabelen waren opgenomen, verklaarde 34% van de variantie in het eindcijfer. (Yu & Jo, 2014) - Significante en positieve relatie tussen het eindcijfer en de totale tijdsduur online. (Firat, 2016) - Geen significante voorspeller in het regressiemodel van You (2016).
Interactie op discussiefora	- Het aantal gelezen discussieberichten en de tijd die hieraan werd besteed, was een significante voorspeller van het eindcijfer. Samen met het aantal bezochte pagina's verklaarde dat 31% in het eindcijfer (Morris et al., 2005) - Geen statistisch significante relatie tussen de bekeken en geplaatste discussieberichten en het eindcijfer. (Davies & Graff, 2005)

	- Het aantal geplaatste discussieberichten, verzonden emails en afgeronde opdrachten verklaarde 33% van de variantie in het eindcijfer. (Macfadyen & Dawson, 2010) - Significante, maar lage correlatie met het eindcijfer. (Conijn, 2017)
Inlogfrequentie	- Positieve correlatie tussen de inlogfrequentie en het eindcijfer. (Macfadyen & Dawson, 2010; Conijn et al., 2017) - Geen significante correlatie tussen de inlogfrequentie en het eindcijfer. (Yu & Jo, 2014; Firat, 2016)
Aanwezigheid	Positieve correlatie tussen de fysieke aanwezigheid in de les en het eindcijfer. (Rodgers, 2001; Kirby & McElroy, 2003; Cohn & Johnson, 2006; Gurung et al., 2010; Akhtar et al, 2017; Gray & Perkens, 2019)

Noot. Er zijn alleen concrete indicaties gegeven van de verklaarde variantie voor zover die ook gerapporteerd zijn in de geraadpleegde artikelen.
Voor de genoemde correlaties geldt $p < .05$

onder andere dat hoe vaker studenten de les bezochten, hoe hoger hun cijfers. Ook Akhtar et al. (2017) vonden een positieve correlatie tussen de fysieke aanwezigheid van studenten in de les en het eindcijfer. In het economie onderwijsdomein hebben Rodgers (2001), Kirby en McElroy (2003) en Cohn en Johnson (2006) hier onderzoek naar gedaan. Uit hun resultaten bleek eveneens dat aanwezigheid in de les een positieve invloed heeft op de studieresultaten en lage aanwezigheid een schadelijk effect op het eindcijfer. Bij hoger onderwijsinstellingen die verplichte aanwezigheid hanteren, en daarmee studenten verplichten de les te bezoeken, werden ook betere studieresultaten gemeten (Credé, Roch, & Kieszczynka, 2010). Fysieke aanwezigheid is dus zeker van invloed op studieresultaten.

1.2 Vraagstellingen en hypothesen

De centrale vraag in dit onderzoek is: "In hoeverre is het haalbaar om bij de opleiding Commerciële Economie learning analytics in te zetten om, op het niveau van de onderwijseenheid, docenten vroegtijdig te informeren over studenten die risico lopen de eenheid niet met succes af te ronden?". Om de centrale vraag te beantwoorden zal eerst worden onderzocht welke relevante data, op basis van de afbakening voor dit onderzoek en de in de learning analytics literatuur gevonden variabelen, bij de opleiding beschikbaar is voor de specifieke onderwijseenheid. Vervolgens zal via een retrospectief onderzoek worden gekeken op welke wijze deze data samenhangt met het wel of niet behalen van de onderwijseenheid. De deelvragen hierbij luiden als volgt:
Deelvraag 1: Voor welke van de voorspellers (variabelen) uit de literatuur zijn er data beschikbaar in de leeromgeving van de opleiding Commerciële Economie?
Deelvraag 2: In hoeverre hebben deze data een voorspellende waarde in deze context?
Deelvraag 3: In hoeverre hebben deze data ook al in een vroeg stadium een voorspellende waarde in deze context?

2. Methode

2.1 Ontwerp

De focus van dit onderzoek was gericht op het identificeren van variabelen die mogelijk in een later stadium of bij vervolgonderzoek als basis kunnen dienen voor de ontwikkeling van een learning analytics datavisualisatie tool voor docenten (Macfadyen & Dawson, 2010). In die zin kan dit onderzoek worden gezien als een eerste stap (analyse fase) van een bredere ontwerpgerichte onderzoeksambitie (EDUCAUSE, 2012). Aan de hand van de afbakening van dit onderzoek en de in de learning analytics literatuur gevonden variabelen werden relevante variabelen in de onderzoekscontext geïdentificeerd, gevolgd door een correlationeel onderzoek – meer in het bijzonder

een retrospectief voorspellend correlationeel onderzoek (Creswell, 2014). Daarbij werd gekeken naar de samenhang tussen variabelen die uit de literatuur als voorspellers naar voren zijn gekomen en het succesvol afronden van de onderwijseenheid CE1: Inleiding Commerciële Economie, van de opleiding CE. Ook werd er specifiek gekeken naar de data van de eerste drie lesweken, omdat het in dit onderzoek gaat om vroegtijdige signalering en de eerste deeltoets van het vak op de maandag van de vierde lesweek werd afgenomen. De afhankelijke variabele bij dit onderzoek was het wel of niet behalen van de onderwijseenheid. Door middel van een logistische regressie analyse kon de relatieve voorspellende waarde van de verschillende variabelen worden vastgesteld. Het onderzoek baseerde zich volledig op bestaande data, die in diverse onderwijssystemen waren opgeslagen. Er was dus sprake van secundaire data-analyse: De onderzoeker heeft de data niet actief zelf verzameld door middel van bijvoorbeeld een enquête.

2.2 Participanten

Het onderzoek werd uitgevoerd met de data van eerstejaars studenten van de HvA opleiding CE ingestroomd in de module CE1: Inleiding Commerciële Economie in het collegejaar 2019-2020. Bij de start van de module waren dat 817 studenten, verdeeld over 28 klassen. Voor het onderzoek zijn alleen de data van de klassen gebruikt waarbij de aanwezigheidsregistratie consequent was uitgevoerd. Dit leverde data op van 12 klassen, 375 studenten in de leeftijd van 18 tot en met 30 jaar. Het combineren van de aanwezigheidsdata met de data uit de DLO en ‘opschonen’ van de dataset leverde een verdere reductie op naar 349 studenten.

2.3 Materialen

Voor de secundaire data-analyse werd gebruik gemaakt van data uit drie verschillende onderwijssystemen, zoals hieronder verder toegelicht.

2.3.1 De digitale leeromgeving

Voor elke onderwijsmodule is er een online pagina ingericht in de DLO Brightspace. Zo ook voor de module CE1. In de online omgeving zet de docent de verschillende leermaterialen klaar en eventuele leeractiviteiten om het leren te ondersteunen. De studenten kunnen te alle tijden de DLO en leermaterialen raadplegen. Daarbij wordt automatisch gebruiksdata van de studenten verzameld en opgeslagen in de DLO. Als een student bijvoorbeeld inlogt in de DLO om de PowerPoint presentatie van de les te bekijken, een opdracht te lezen dan wel in te leveren of aanvullend materiaal te bestuderen, dan is in de DLO te zien dat hij/zij de pagina heeft bezocht en hoeveel tijd hieraan besteed is. Het totaal aantal DLO pagina’s bedraagt 24 voor de gehele onderwijsperiode. Daarvan hebben 10 pagina’s betrekking op de eerste drie lesweken.

2.3.2 De aanwezigheidsregistratietool

De applicatie Academy Attendance wordt gebruikt bij de opleiding voor het registreren van de fysieke aanwezigheid van studenten in de les. Dit gebeurt echter niet automatisch. Docenten registreren per klas de aanwezigheid van individuele studenten handmatig in de applicatie of presenteren een QR-code in de les, die de studenten zelf kunnen scannen. Op het moment dat de student de QR-code scant, wordt de aanwezigheid geregistreerd. Dit vereist dus een actieve handeling van de docenten en/of student en gebeurt als gevolg daarvan niet altijd even consequent. In de applicatie is per student te zien welke lessen hij/zij heeft bijgewoond en als er voor één of meerdere lessen geen aanwezigheid is geregistreerd. Voor dit onderzoek is gewerkt met data van klassen (c.q. studenten) die minimaal 14 lessen hadden gevolgd en daarvan bij maximaal 2 lessen de aanwezigheid niet was geregistreerd. Er is uitgegaan van 14 lessen omdat de module, zoals ingericht in de DLO, zeven lesweken beslaat met twee lessen per week. Dit leverde data op van 12 klassen.

2.3.3 Het Studenten Informatie Systeem

De afhankelijke variabele, het wel of niet behalen van de onderwijsmodule, is een categorische variabele. Hiervoor wordt gebruik gemaakt van het eindresultaat (cijfer) zoals geregistreerd in SIS. De module is succesvol afgerond als het gewogen eindcijfer onafgerond ten minste een 5,5 is. De studenten met het eindcijfer 5,45 of lager hebben de module niet behaald. Daarnaast worden in het SIS ook algemene gegevens van de student geregistreerd zoals geslacht.

2.4 Procedure

Omdat er voor dit onderzoek gebruik wordt gemaakt van secundaire data – die gedurende de onderwijsperiode wordt gegenereerd en opgeslagen in diverse onderwijssystemen – was het niet nodig om de studenten actief te benaderen en toestemming te vragen voor het onderzoek. Dit conform de richtlijnen van de ethische commissie van de Open Universiteit (cETO, 2016) en de ethische richtlijnen van de American Psychological Association (American Psychological Association, 2002). Het gebruik van deze data voor secundaire data-analyse gebeurt met schriftelijke toestemming van de opleidingsmanager van CE en de informatiemanager annex contactpersoon Algemene Verordening Gegevensbescherming (AVG) van de HvA Faculteit Business en Economie.

De onderzoeker heeft in nauwe samenwerking met de AVG-contactpersoon de secundaire dataset gegenereerd en volledig geanonimiseerd. De dataset is tot stand gekomen via de volgende procedure: De kwaliteit van de aanwezigheidsdata is geanalyseerd en op basis van de genoemde criteria voor het aantal gevolgde lessen is er een selectie gemaakt van de klassen. Vervolgens is voor de studenten uit de betreffende klassen een dataset gecreëerd met aanwezigheidsdata, aangevuld met data uit de DLO. De data uit de DLO zijn ook kritisch geanalyseerd. Aan deze dataset is vervolgens de data uit het SIS toegevoegd en is het bestand volledig geanonimiseerd, zodat de gegevens niet herleidbaar zijn naar individuele studenten. De analyse en interpretatie van de resultaten heeft de

onderzoeker ook voorgelegd aan en besproken met enkele collega's van de afdeling Onderwijs en Onderzoek en de afstudeerbegeleider.

2.5 Data-analyse

Eerst werd onderzocht voor welke potentiële voorspellers uit de literatuur (tabel 1) er data in de huidige onderzoekscontext beschikbaar waren. Vervolgens zijn de relevante variabelen voor dit onderzoek geoperationaliseerd en zijn de bijbehorende data verzameld. Het volledig geanonimiseerde databestand werd voor analyse geïmporteerd in SPSS versie 25. Via data-exploratie werd de data eerst gecontroleerd op kwaliteit. Als gevolg van een technische error (*system time out)* in de DLO is bij sommige studenten geen tijdsduur geregistreerd ondanks dat zij wel bepaalde content bekeken hebben. De tijdsduur van deze cases werd gerapporteerd als missing value. Daarnaast bevatte een van de DLO pagina's alleen een link naar een quiz in een externe applicatie. Er was alleen informatie beschikbaar over de studenten die de link hadden aangeklikt, en niet of ze überhaupt de quiz hadden gestart, dan wel gemaakt en hoeveel tijd ze hadden besteed aan het maken van de quiz. Deze data is daarom verwijderd uit de dataset (niet meegenomen in het totaal aantal bezochte pagina's).

Ook werd er een vooranalyse gedaan om de niet-dichotome variabelen aanwezigheid en de tijdsduur per bekeken DLO pagina's te controleren op outliers. Hiertoe werden per variabele de z-scores berekend. Field (2013) stelt dat alle cases met een z-waarde groter dan 3.29 tot een extreme outlier gerekend kunnen worden. Bij de tijdsduur-variabelen betrof het extremen in de tijdsduur die studenten hadden besteed aan het bestuderen van de informatie in de DLO. De studenten hebben dus wel degelijk tijd hieraan besteed – vermoedelijk ook meer dan gemiddeld – maar zijn wellicht tussendoor ook met andere zaken bezig geweest. Omdat deze extreme outliers voor bias in de data kunnen zorgen, is daarom gekozen om *winsorizing* toe te passen. Daarbij zijn de outliers vervangen door de eerstvolgende hoge score die niet als outlier is aangemerkt (Field, 2013). Deze werkwijze geniet de voorkeur boven het eenvoudig verwijderen van de outliers, omdat het volledig verwijderen van de outliers een verlies aan informatie betekent die de analyseresultaten behoorlijk kan vertekenen. Bij de gemeten aanwezigheid werden ook vier outliers gedetecteerd in de data. Op basis van een nadere analyse van de data van deze respondenten bleek het in drie gevallen te gaan om administratie/systeemfouten; betreffende participanten behoorden niet tot de doelpopulatie en zijn verwijderd uit de dataset. De vierde outlier betrof een reguliere student die bij geen van de lessen aanwezig was geweest maar wel het tentamen had gemaakt. Omdat dit een reëel scenario is, werd ervoor gekozen om deze outlier niet te verwijderen. Vervolgens zijn de beschrijvende statistieken (gemiddelden en standaarddeviaties) voor de afhankelijke en onafhankelijke variabelen berekend.

De samenhang tussen variabelen die uit de literatuur als voorspellers naar voren zijn gekomen en het succesvol afronden van de onderwijseenheid werd onderzocht via een binaire logistische regressieanalyse. Er werd gekozen voor een binaire logistische regressieanalyse omdat de uitkomstvariabele – het wel of niet behalen van de onderwijseenheid – een dichotome variabele is.

Eerst werd gecontroleerd of aan de data-assumpties voor deze analyse was voldaan. Daarbij werden de residuen gecontroleerd om na te gaan hoe goed het model past bij de data: zowel de Cook's distance en DF Beta waarden zijn kleiner dan 1 en de *standardized residuals* vallen binnen het bereik van de gestelde grenswaarden ± 1,96 (Field, 2013). Ook werd de assumptie gecontroleerd die stelt dat de predictorvariabelen lineair gerelateerd moeten zijn aan de log van de uitkomstvariabele. Hiervoor werd gekeken naar de interacties tussen predictorvariabelen en de log van de predictorvariabele; deze interactie-effecten waren niet significant en de assumptie van lineariteit is dus niet geschonden. Daarnaast zijn er geen aanwijzingen gevonden dat de assumptie van multicollineariteit is geschonden. De VIF-waarden zijn kleiner dan de grenswaarde van 10 en de tolerantie is groter dan de grenswaarde 0,1 (Field, 2013). Er is dus voldaan aan de assumpties voor het uitvoeren van een binaire logistische regressieanalyse.

Bij het uitvoeren van de analyse werd de algemene procedure die Field (2013) voorschrijft, aangehouden: een hiërarchische regressie analyse (methode: ENTER) werd uitgevoerd waarbij de verschillende mogelijke modellen werden vergelijken en de *'most parsimonious'* werd gekozen. Er werden vervolgens twee analyses uitgevoerd om deelvragen 2 en 3 te beantwoorden. Bij de eerste analyse werd gekeken naar de variabelen totale aanwezigheid, totaal aantal DLO pagina's bestudeerd en de totale tijdsduur die hieraan besteed werd, over de gehele onderwijsperiode - en het wel of niet behalen van de onderwijseenheid. Er werd gekozen voor deze volgorde omdat uit eerdere onderzoeken consequent naar voren kwam dat er een positief verband is tussen aanwezigheid en studiesucces; voor de overige twee variabelen zijn de resultaten uit eerder onderzoek verschillend. In de tweede analyse werd met dezelfde data, maar dan specifiek van de eerste drie lesweken, onderzocht of er een verband was met het wel of niet behalen van de onderwijseenheid. Bij de analyses werd aan de hand van de beta waarden gekeken of en welke variabelen significante voorspellers waren en werd de Exp(B) oftewel de odds ratio geïnterpreteerd. Als de odds ratio groter dan 1 is, dan is de kans ook groter dat – als de waarde van de predictor toeneemt – het vak wordt behaald. Daarnaast werd gekeken hoe goed de predictorvariabelen het wel of niet behalen van vak voorspellen. In dit onderzoek werd hiervoor de Nagelkerke's R^2 waarde gebruikt; voor logistische regressies analyses is er in de literatuur nog geen consensus over welke maat hiervoor het beste gebruikt kan worden (Sieben & Linssen, 2009).

3. Resultaten

In dit hoofdstuk worden de resultaten beschreven voor het beantwoorden van de drie deelvragen. Eerst worden de potentiële voorstellers gepresenteerd voor welke er data beschikbaar zijn in de onderzoekscontext. Vervolgens worden de resultaten van de statistische analyses, om de voorspellende waarde vast te stellen, weergegeven.

3.1 Potentiële voorspellers in de onderzoekscontext: beschikbare data

Voor het beantwoorden van deelvraag 1 is vastgesteld voor welke potentiële voorspellers uit de literatuur er data beschikbaar waren in de geselecteerde onderwijssystemen van de opleiding CE. In Tabel 2 zijn de variabelen uit Tabel 1 overgenomen en waar mogelijk aangevuld met data die daarvoor beschikbaar waren in de onderzoekscontext. Op basis van de beschikbare data zijn de onderzoeksvariabelen vastgesteld en geoperationaliseerd. In de tabel is ook de afhankelijke variabele opgenomen.

3.2 Voorspellende waarde binnen de onderzoekscontext

Voor het beantwoorden van deelvraag 2 en 3 zijn de beschrijvende statistieken gepresenteerd evenals de statistische analyses die betrekking hebben op de voorspellende waarde van de data.

3.2.1 Beschrijvende statistieken

Van de 349 studenten had 34.7% het vak niet behaald. In tabel 3 zijn de gemiddelden en standaarddeviaties van de voor het onderzoek geselecteerde variabelen over gehele onderwijsperiode gepresenteerd (totaal variabelen) evenals de beschrijvende statistieken voor de eerste drie lesweken (vroegtijdige voorspellers). Van de data van de eerste drie lesweken was in de DLO echter niet te achterhalen of de huidige aantallen ook nog latere bezoeken omvatten, omdat alleen de totaalstand wordt bijgehouden en geen historische data. Studenten hebben de DLO pagina’s van week 1 tot en met 3 niet alleen gedurende de drie weken geraadpleegd, maar mogelijk ook daarna.

Tabel 2

De voor het onderzoek relevante data in de CE onderwijssystemen

Variabele	**Beschikbare data in onderzoekscontext**	**Operationalisatie (afgeleide data)**	**Variabele naam**	**Hoe gemeten**
Aantal DLO pagina's bezocht	Aantal bezoeken per DLO pagina	Totaal aantal **unieke** DLO pagina's bezocht, over de gehele onderwijsperiode (0-24)	pages_totaalbezocht	Data verzameld in DLO logboek
		Totaal aantal **unieke** DLO pagina's bezocht van lesweek 1 t/m 3 (0-10)	pagesweek1-3_totaalbezocht	
Tijdsduur online in de DLO	Duur paginabezoek	Totale tijdsduur DLO pagina's bekeken (h:m:s)	tijd_totaal	Data verzameld in DLO logboek: daarbij is de totale tijdsduur tussen het in- en uitloggen gemeten
		Totale tijdsduur DLO pagina's week 1 t/m 3 bekeken (h:m:s)	tijd_week1-3	
Interactie op discussiefora	X	X	X	X
Inlogfrequentie	X	X	X	X
Aanwezigheid	Aanwezigheid per les	Proportie van het totaal aantal lessen aanwezig (%) Proportie van het totaal aantal lessen aanwezig (%) in week 1 t/m 3	aanwezigheid aanwezigheid_week1-3	Aanwezigheid in de les geregistreerd
Afhankelijke variabele: Module afgerond	Cijfers en studiepunten in SIS	Studiepunten behaald op basis van een voldoende (≥ 5,5)	vak_behaald	Resultaten in SIS geregistreerd

Tabel 3

Beschrijvende statistieken aanwezigheid, aantal unieke DLO pagina's bezocht en tijdsduur

	Module niet behaald (*n* = 121)		Module wel behaald (*n* = 228)	
	M	*SD*	*M*	*SD*
Proportie aanwezig totaal	.61	.22	.72	.16
Totaal aantal unieke DLO pagina's bezocht (0-24)	11.76	5.77	14.72	5.37
Totale tijdsduur DLO pagina's bekeken	2:50:14	3:16:57	4:15:17	3:54:42
Proportie aanwezig in week 1 t/m 3	.83	.23	.92	.13
Totaal aantal unieke week 1 t/m 3 DLO pagina's bezocht (0-10)	6.02	2.63	6.79	2.38
Totale tijdsduur week 1 t/m 3 DLO pagina's bekeken	1:22:31	1:32:45	1:48:50	1:39:27

3.2.2 Voorspellende waarde totaal variabelen

Voor het analyseren van de voorspellende waarde van de totaal variabelen aanwezigheid, aantal unieke pagina's bezocht en tijdsduur werd een hiërarchische regressieanalyse uitgevoerd (Field, 2013). Hiervoor werden allereerst de variabelen Aanwezigheid, Pages_totaalbezocht en Tijd_totaal, bloksgewijs toegevoegd om het best passende model te kiezen (Bijlage 1). De uitkomstvariabele was daarbij de dichotome categorische variabele: vak behaald. Het model met de voorspellers Aanwezigheid en Pages_totaalbezocht bleek een significant model voor de data, $\chi^2(2) = 34.02$, $p = .00$. Het toevoegen van de predictor variabele Tijd_totaal aan het model was niet significant $\chi^2(3) = 0.1$, $p = .75$. Model 2 leverde dus een significant beter model op ten opzichte van model 3. Voor dit model werden de assumpties gecontroleerd. Aan de assumptie voor lineariteit is voldaan ($p > .05$), evenals de assumptie voor multicollineariteit (VIF = 1.28, tolerantie = 0.78). Ook zijn de residuen gecontroleerd om na te gaan hoe goed het model past bij de data: zowel de Cook's distance en DF Beta waarden zijn kleiner dan 1. Van de *standardized residuals* is slecht 3% boven de grenswaarde van ±1.96. Er zijn dus geen extreme outliers in de data die een effect hebben op het model.

De analyse werd uitgevoerd met de predictorvariabelen Aanwezigheid en Pages_totaalbezocht (Bijlage 2). In tabel 4 is te zien dat het model over het geheel genomen in 69.6% van de gevallen een correcte voorspelling maakte, maar van de groep die het vak niet behaald werd dit bij slechts 27.3% correct voorspeld. Uit de logistische regressieanalyse bleek dat beide predictoren significante voorspellers zijn voor het wel of niet behalen van de onderwijseenheid, maar de variabele

Pages_totaalbezocht had een relatief lage odds ratio. Ook bepaalde het gehele model slechts 13% van de variantie (Nagelkerke) in het wel of niet behalen van de onderwijseenheid. In tabel 5 worden de resultaten van de analyse weergegeven.

Tabel 4

Classificatie van studenten: module wel of niet behaald (n = 349)

Geobserveerd	**Voorspeld**		
	Module niet behaald	Module behaald	Percentage correct
Module niet behaald	33	88	27.3
Module behaald	18	210	92.1
Overall Percentage			69.6

Tabel 5

Geschatte parameters van het logistische regressiemodel dat voorspelt of een student de module behaalt (n = 349)

			95% betrouwbaarheidinterval voor Exp(B)		
	b	*p*	*ondergrens*	*Exp(B)*	*bovengrens*
Aanwezigheid	2.39	.001	2.81	10.94	42.56
Pages_totaalbezocht	.06	.011	1.01	1.06	1.11
Constant	-1.76	.000		.17	

De variabelen Aanwezigheid en Pages_totaalbezocht zijn beide significante predictoren in dit model, $p < .05$. Voor de variabele aanwezigheid geldt dat als deze toeneemt, de kans op het behalen van het vak ook toenoemt (*Exp(B)* = 10.94). Voor de predictorvariabele Pages_totaalbezocht is de *Exp(B)* = 1.06. De waarde 1 fungeert als drempelwaarde: bij een *Exp(B)*-waarde groter dan 1 neemt de kans toe en bij een waarde kleiner dan 1 neemt de kans af (Field, 2013). Het betrouwbaarheidsinterval van deze voorspeller bevat ook de waarde 1 en dus kan er niet met zekerheid gesteld worden dat als het aantal bezochte pagina's toeneemt, de kans ook groter wordt dat een student het vak behaald.

3.2.3 Vroegtijdige voorspellers

Om te onderzoeken of de voorspellende variabelen ook al vroegtijdig een signaalfunctie kunnen vervullen, werd een logistische regressieanalyse uitgevoerd met data die betrekking hebben op de eerste drie lesweken. Ook hier werd de analyse uitgevoerd conform de stappen die Field (2013) beschrijft. De variabelen Aanwezigheid week 1-3, Pagesweek1-3_totaalbezocht en Tijd_week1-3 werden bloksgewijs toegevoegd om het best passende model te bepalen (Bijlage 3). Alleen het eerste model met de voorspeller Aanwezigheid week 1-3 bleek een significant model voor de data, $\chi^2(1) = 22.38$, $p = .00$. Het toevoegen van de overige predictorvariabelen leverde geen significant betere modellen op, $p > .05$. Aan de assumptie voor lineariteit is voldaan ($p > .05$) evenals de assumptie voor

multicollineariteit (VIF = 1.00, tolerantie = 1.00). Ook zijn de residuen gecontroleerd om na te gaan hoe goed het model past bij de data: zowel de Cook's distance en DF Beta waarden zijn kleiner dan 1 en de *standardized residuals* liggen allemaal binnen het bereik van de grenswaarden van ±1.96. Er zijn dus geen extreme outliers in de data die een effect hebben op het model.

Op basis hiervan werd de analyse uitgevoerd met de predictorvariabele Aanwezigheid (Bijlage 4). In tabel 6 is te zien dat het model in 68.8% van de gevallen een correcte voorspelling maakte, maar voor de groep die het vak niet behaald werd dit bij slechts 27.3% correct voorspeld. Uit de logistische regressieanalyse blijkt dat aanwezigheid in de eerste drie weken een significante voorspeller is voor het wel of niet behalen van de onderwijseenheid. Echter, het model bepaalt slechts 8.6% van de variantie (Nagelkerke) in het wel of niet behalen van de onderwijseenheid. In tabel 7 worden de resultaten van de analyse weergegeven.

Tabel 6

Classificatie van studenten: module wel of niet behaald (n = 349)

Geobserveerd	**Voorspeld**		
	Module niet behaald	Module behaald	Percentage correct
Module niet behaald	33	88	27.3
Module behaald	21	207	90.8
Overall Percentage			68.8

Tabel 7

Geschatte parameters van het logistische regressiemodel dat op basis van de data van lesweken 1 t/m 3 voorspelt of een student de module behaalt (n = 349)

					95% betrouwbaarheidinterval voor Exp(B)
	b	*p*	*ondergrens*	*Exp(B)*	*bovengrens*
Aanwezig week 1 tm 3	3.02	.000	5.51	20.40	75.56
Constant	-2.03	.001		.13	

De variabele Aanwezigheid is een significante predictor in dit model, $p < .05$. Voor de variabele aanwezigheid geldt dat als deze toeneemt, de kans op het behalen van het vak ook toeneemt (*Exp(B)* = 20.40).

4. Conclusie en discussie

Het doel van dit onderzoek was om na te gaan in hoeverre het bij de opleiding CE haalbaar is om op het niveau van een onderwijseenheid learning analytics in te zetten om docenten vroegtijdig te informeren over studenten die het risico lopen de module niet met succes af te ronden. Om dit te onderzoeken werd gebruik gemaakt van data uit de digitale leeromgeving en de aanwezigheidsregistratietool van de opleiding. Waar in de literatuur al veel bekend is over de positieve effecten van aanwezigheid op studieresultaten en studiesucces, zijn in de wetenschappelijke learning analytics literatuur geen eenduidige voorspellers uit de digitale leeromgeving gerapporteerd en onderzoek in de eigen context werd aanbevolen.

4.1 Conclusie

Voor dit onderzoek werd eerst nagegaan voor welke van de in de literatuur gevonden voorspellers er data beschikbaar was in de leeromgeving van de opleiding. Daarbij werden de volgende predictoren geoperationaliseerd voor de gehele onderwijsperiode: het totaal aantal bezochte DLO pagina's, de daaraan bestede tijdsduur en de aanwezigheid in de les. Ook werden vroegtijdige voorspellers geoperationaliseerd op basis van de data van de eerste drie lesweken: het totaal aantal bezochte DLO pagina's van de betreffende lesweken, de daaraan bestede tijdsduur en de aanwezigheid in de les. Over de gehele onderwijsperiode genomen bleken alleen de totale aanwezigheid en het totaal aantal bezochte pagina's aan het einde van de onderwijsperiode het wel of niet behalen van de module te voorspellen. Alhoewel het model over het geheel genomen in 69.6% van de gevallen een correcte voorspelling maakte, werd van de groep studenten die het vak niet behaalt slechts in 27.3% van de gevallen een correcte voorspelling gemaakt. Waar het ging om vroegtijdige signalering van deze risicostudenten, bleek dat alleen de aanwezigheid in de eerste drie lesweken een significante voorspeller was. Het aantal bezochte DLO pagina's en de hieraan bestede tijdsduur in de eerste drie lesweken bleken hier geen significante voorspellers. Echter werd ook in dit model slechts voor 27.3% van de studenten die het vak niet behalen een correcte voorspelling gemaakt. Alhoewel met dit model en deze voorspellers de opleiding dus wel een deel van de risicostudenten via learning analytics in kaart zou kunnen brengen en een interventie zou kunnen plegen, is het wenselijk dat het model eerst verder geoptimaliseerd wordt om betere voorspellingen te doen. Voorspellingen doen op basis van een model dat ongeveer 72% van de risicostudenten niet correct voorspelt – terwijl het juist de bedoeling is om zoveel mogelijk deze groep tijdig te signaleren – is niet eerlijk naar deze groep en daarmee wellicht ook niet helemaal ethisch. Ook bleek uit de analyses dat een kleine groep studenten onterecht wordt gecategoriseerd in de groep die het vak niet zal behalen. Op een onjuiste manier een interventie plegen kan deze studenten mogelijk onzeker maken, of voor studenten die al twijfels hadden over het kunnen behalen van het vak een self-fullfilling prophecy te creëren. Het huidige model is daarmee niet

voldoende accuraat en biedt dus onvoldoende handvatten om op het niveau van onderwijseenheid docenten vroegtijdig te informeren over studenten die het risico lopen de module niet met succes af te ronden.

4.2 Discussie

In dit onderzoek bleek de variabele aanwezigheid de relatief sterkste voorspeller van het wel of niet behalen van de module, ook als vroegtijdige indicator. Dit sluit aan bij de verwachting uit eerdere onderzoeken naar het verband tussen aanwezigheid in de les en de studieresultaten. Daarbij werd gesteld dat aanwezigheid in de les een positieve invloed heeft op studieresultaten en beperkte lage aanwezigheid dus een schadelijk effect heeft op het eindcijfer (Rodgers, 2001; Kirby & McElroy, 2003; Cohn & Johnson, 2006; Grunung et al., 2010; Akthar et al., 2017). Ook uit het onderzoek van Gray en Perkins (2019) bleek aanwezigheid een sterke voorspeller van het eindcijfer. Zij ontwikkelden op basis van aanwezigheidsdata een model dat reeds in een heel vroeg stadium met 97% betrouwbaarheid de studenten kon signaleren die het vak niet zouden behalen. Alhoewel de odds ratio en het betrouwbaarheidsinterval in dit onderzoek wel een indicatie geven dat de variabele aanwezigheid een betrouwbare voorspeller is, zien wij echter dat het model waarin deze variabele is opgenomen toch onvoldoende in staat is goed de risicostudenten te voorspellen; dit gebeurt in slechts 27.3% van alle risicogevallen. Van het grootste deel van de studenten, die het vak niet behaalt, wordt dus voorspeld dat ze de module wel succesvol zullen afronden.

Verder bleek in dit onderzoek alleen het totaal aantal bezochte pagina's tijdens de gehele onderwijsperiode ook een significante voorspeller in het model, samen met de aanwezigheid. Het aantal bezochte pagina's speelde verder geen rol bij het vroegtijdig signaleren van studenten die de module niet behalen. Eerdere onderzoeksresultaten zijn niet volledig eenduidig wat betreft de voorspellende waarde van het aantal bezochte DLO pagina's. Zo bleek deze predictor – in een model met meerdere voorspellers – in het onderzoek van Morris et al. (2015) wel een significante voorspeller van het eindcijfer en uit onderzoek van Conijn et al. (2017) bleek het zelf één van de meest stabiele voorspellers. Daarentegen bleek uit het onderzoek van Macfadyen en Dawson (2010) echter geen voorspellende waarde. In het huidige onderzoek liggen de odds ratio en het betrouwbaardheidsinterval van deze voorspeller allemaal rond de waarde 1. Dat maakt het geen betrouwbare voorspeller. Immers, bij een odds ratio van 1 kun je niet met zekerheid zeggen dat als een student meer DLO pagina's bekijkt de kans ook echt groter is dat het vak wordt behaald.

Naast de aanwezigheid en het aantal bezochte pagina's werd ook de predictor tijdsduur online toegevoegd aan het model. Alhoewel uit eerdere onderzoeksresultaten een positieve correlatie bleek tussen de tijdsduur online in de DLO en het eindcijfer (Macfadyen & Dawson, 2010; Rybov, 2012; Yo & Jo, 2014: Firat, 2016) bleek deze predictor zowel na de eerste drie lesweken als aan het einde van de onderwijsperiode, geen significante bijdrage te leveren aan het model en het voorspellen van het (niet) succesvol afronden van de onderwijseenheid. Ook hier zijn eerdere onderzoeksresultaten niet

eenduidig. In het onderzoek van You (2016) bleek de tijdsduur online ook geen significante voorspeller, maar in het onderzoek van Firat (2016) juist weer wel. Dat de tijdsduur online niet significant bijdraagt aan het model houdt mogelijk verband met de argumentatie van Lowes, Lin en Kinghorn (2015) dat de tijdsduur online niet perse hetzelfde is als de tijd die een student ook daadwerkelijk aan een taak heeft besteed. Sommige studenten loggen bijvoorbeeld alleen in in de DLO om op te halen wat zij nodig hebben en werken dan offline of buiten de leeromgeving verder en weer andere studenten blijven ingelogd, maar zijn daar niet perse aan het werk (Lowes et al., 2015).

De resultaten uit het huidige onderzoek bevestigen daarmee het algemene beeld uit de learning analytics literatuur wat betreft voorspellers uit de DLO. Resultaten uit eerder onderzoek zijn niet of nauwelijks te generaliseren naar andere contexten en ook in dit onderzoek roepen de resultaten de vraag op of je überhaupt kunt komen tot een set algemene predictoren, of dat alle onderwijsinstellingen, onderwijseenheden en studenten zo divers zijn dat een eigen model om voorspellingen te doen, nodig is (Conijn et al., 2017; Larrabee Sønderlund, Hughes, & Smith, 2019). Als het daarnaast gaat om het voorspellen van studenten die de onderwijseenheid niet zullen behalen, is de accuraatheid van het huidige voorspellingsmodel laag. De lage accuraatheid van het model komt mogelijk ook door de geselecteerde data en inrichting van de DLO. Het is niet eenvoudig om betekenisvolle indicatoren, die iets zeggen over het leerproces, te destilleren uit datasets en om goede voorspellingen te kunnen doen is het belangrijk om deze indicatoren ook te koppelen aan theoretische concepten over bijvoorbeeld het leergedrag van studenten (Greller en Draschler, 2012; You, 2016).

Beperkingen en vervolgonderzoek

Het huidige onderzoek kent een aantal beperkingen. Ten eerste is er voor dit onderzoek gewerkt met data uit de DLO die vooral gebruikt werd ter ondersteuning van traditioneel face-to-face onderwijs. De leeromgeving was niet weldoordacht en bewust ingericht met leeractiviteiten die aansluiten op een bepaald onderwijskundig ontwerp of *blended* onderwijsconcept. De DLO werd vooral ondersteunend gebruikt voor het opslaan en delen van PowerPoints van de les, additioneel leesmateriaal en een enkele inlevermap voor een opdracht. Als de DLO wordt ingericht aan de hand van een bepaald onderwijskundig concept met weldoordachte leeractiviteiten en leermaterialen en studietaken fijnmaziger worden gepresenteerd, kan er mogelijk ook veel waardevollere data voor learning analytics doeleinden worden gegeneerd.

Daarnaast zijn in het huidige onderzoek de predictoren geselecteerd op basis van de data die beschikbaar en eenvoudig te verzamelen waren in de onderwijssystemen. Het onderzoek is daarmee vooral data-gedreven en niet betekenisvol ingebed in een onderwijs- of leertheorie. Het risico daarmee is ook dat het accent vooral komt te liggen op de vraag of een student de module succesvol afrondt zonder dat er met zekerheid gesteld kan worden dat er ook daadwerkelijk sprake is van leren. Clow (2012) spreekt hier van: "*optimising to a metric that does not reflect what is desired as an outcome. Learning analytics should generate metrics that relate to what is valued in the learning process."*(p.

137). Voor vervolgonderzoek is het noodzakelijk om vast te stellen hoe theoretische concepten geoperationaliseerd en gemeten kunnen worden in de onderwijssystemen. Dit kan leiden tot betere predictoren en interpretatie van de resultaten.

Ook was de beschikbaarheid van data en daarmee van de onderzochte predictoren voor dit onderzoek beperkt. Dyckhoff, Lukarov, Muslim, Chatti en Schroeder (2013) stellen dat data die door studenten is gegenereerd in de DLO niet voldoende is om een holistisch beeld te creëren van het leerproces van een student. Uit de modellen blijkt ook een lage variantie: 13% respectievelijk 8.6% van de variantie (Nagelkerke) in het wel of niet behalen van de onderwijseenheid wordt door de predictoren verklaard. Voor vervolgonderzoek kunnen mogelijk meerdere databronnen gecombineerd worden, bijvoorbeeld informatie over de vooropleiding van de student, studentkarakteristieken of data die nu niet eenvoudig te raadplegen zijn. Ook zijn in het huidige onderzoek de resultaten van de tussentijdse deeltoets in week 4 niet meegenomen als voorspeller, maar uit de literatuur blijkt dat de scores op tussentijdse toetsen een behoorlijke voorspellende waarde hebben (Van Berkel, Jansen, & Bax, 2012). Verder bleek bij het selecteren en controleren van de data voor dit onderzoek, dat deze niet één op één gebruikt konden worden. Aanwezigheidsregistratie gebeurde bijvoorbeeld niet overal even betrouwbaar en consequent en ook in de DLO was kritische analyse van de data nodig. Daardoor is uiteindelijk gewerkt met minder data. Ook is in die onderzoek slechts van één onderwijseenheid data verzameld en voor vervolgonderzoek kan dit uitgebreid worden. Met meerdere databronnen en predictoren is mogelijk te komen tot een beter en accurater voorspellingsmodel. Voor vervolgonderzoek is het daarbij ook interessant om te onderzoeken wanneer – bijvoorbeeld bij welke drempelwaarde – een learning analytics voorspellingsmodel voldoende accuraat en betrouwbaar is. Echter moeten wij ons wel realiseren dat een model dat 100% correcte voorspellingen maakt, na interventie niet automatisch ook zal leiden tot een slagingspercentage van 100%. Er zullen wel degelijk een aantal studenten zijn voor wie het niveau of de inhoud van het vak te complex is.

Als de opleiding learning analytics wil in zetten om docenten vroegtijdig te informeren over studenten die het risico lopen een module niet met succes af te ronden, zal de opleiding op basis van bovengenoemde discussiepunten samen met de onderwijskundige eerst kritisch moeten kijken naar de (kwaliteit van de) data die gebruikt wordt voor learning analytics en hoe variabelen te operationaliseren aan de hand van theoretische concepten. Dit kan zich vertalen naar het maken van goede afspraken binnen de docententeams van de onderwijseenheden, samen met de onderwijskundige, over de wijze waarop de DLO wordt ingezet en ingericht. Daarnaast is het ook belangrijk dat de aanwezigheidsregistratie consequent gebeurt, wil de opleiding dit voor voorspellende learning analytics doeleinden inzetten. Dit gebeurt nu nog niet altijd, met als gevolg onvolledige en onbetrouwbare data. Dat kan mogelijk ook verklaren waarom in dit onderzoek de predictor aanwezigheid zo een kleine rol speelt in het voorspellen van het wel of niet behalen van de onderwijseenheid: Er is in dit onderzoek slechts van één onderwijseenheid aanwezigheidsdata verzameld en door de onvolledigheid hiervan kon slechts een deel voor analyse gebruikt worden.

Dit onderzoek werd gezien als eerste, analyserende stap van een bredere ontwerpgerichte onderzoeksambitie om variabelen te identificeren die mogelijk in een later stadium als basis kunnen dienen voor de ontwikkeling van een learning analytics datavisualisatietool voor docenten. Ondanks bovengenoemde beperkingen biedt het interessante perspectieven en bevindingen om verder binnen de opleiding, maar ook binnen de faculteit, de dialoog met elkaar aan te gaan over learning analytics en de mogelijkheden die learning analytics ons biedt, verder te verkennen.

Referenties

Akhtar, S., Warburton, S., & Xu, W. (2017). The use of an online learning and teaching system for monitoring computer aided design student participation and predicting student success. *International Journal of Technology and Design Education, 27*(2), 251-270. doi:10.1007/s10798-015-9346-8

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *The American Psychologist, 57*(12), 1060-1073. doi:10.1037/0003-066X.57.12.1060

Archer, E., & Prinsloo, P. (2020). Speaking the unspoken in learning analytics: troubling the defaults. *Assessment & Evaluation in Higher Education*, *45*(6), 888-900. doi:10.1080/02602938.2019.1694863

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 267-270). doi:10.1145/2330601.2330666

Bienkowski, M., Feng, M., & Means, B. (2012, oktober). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*. Geraadpleegd van https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf

Brennan, A., Sharma, A., & Munguia, P. (2019). Diversity of Online Behaviours Associated with Physical Attendance in Lectures. *Journal of Learning Analytics*, *6*(1), 34-53. doi:10.18608/jla.2019.61.3

cETO (2016). *Onderzoek secundaire data*. Geraadpleegd van https://www.ou.nl/documents/40554/361215/3_Onderzoek_secundaire_data_14122016.pdf/b6e0c79a-7ce8-61a8-8d0b-b297f38f61e2

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning, 4*(5/6), 318-331. doi:10.1504/ijtel.2012.051815

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 134-138). https://doi.org/10.1145/2330601.2330636

Coates, H. (2005). The value of student engagement for higher education quality assurance. *Quality in Higher Education, 11*(1), 25-36. doi:10.1080/13538320500074915

Cohn, E., & Johnson, E. (2006). Class attendance and performance in principles of economics. *Education Economics, 14*(2), 211-233. doi:10.1080/09645290600622954

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting student performance from LMS data: A comparison of 17 blended courses using moodle LMS. *IEEE Transactions on Learning Technologies, 10*(1), 17-29. doi:10.1109/TLT.2016.2616312

Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research, 80*(2), 272-295. doi:10.3102/0034654310362998

Creswell, J. W. (2014). *Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research*. Essex, England: Pearson.

Davies, J., & Graff, M. (2005). Performance in e-learning: Online participation and student grades. *British Journal of Educational Technology, 36*(4), 657-663. doi:10.1111/j.1467-8535.2005.00542.x

Delnoij, L. E. C., Dirkx, K. J. H., Janssen, J. P. W., & Martens, R. L. (2020). Predicting and resolving non-completion in higher (online) education – A literature review. *Educational Research Review, 29*, 100313. doi:10.1016/j.edurev.2020.100313

Dietz-Uhler, B., & Hurn, J. E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning, 12*(1), 17. Geraadpleegd van http://www.ncolr.org/jiol/issues/pdf/12.1.2.pdf

Drachsler, H., & Greller, W. (2012). The pulse of learning analytics understandings and expectations from the stakeholders. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 120-129). doi:10.1145/2330601.2330634

Dyckhoff, A. L., Lukarov, V., Muslim, A., Chatti, M. A., & Schroeder, U. (2013). Supporting action research with learning analytics. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 220-229). Geraadpleegd van https://dl.acm.org/doi/abs/10.1145/2460296.2460340

EDUCAUSE (2012). *7 Things you should know about educational design research*. Geraadpleegd van https://library.educause.edu/-/media/files/library/2012/8/eli7087-pdf.pdf

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. SAGE Publications.

Firat, M. (2016). Determining the effects of LMS learning behaviors on academic achievement in a learning analytic perspective. *Journal of Information Technology Education: Research, 15*, 75-87. Geraadpleegd van http://www.jite.informingscience.org/documents/Vol15/JITEv15ResearchP075-087Firat1928.pdf

Gray, C. C., & Perkins, D. (2019). Utilizing early engagement and machine learning to predict student outcomes. *Computers and Education, 131*, 22-32. doi:10.1016/j.compedu.2018.12.006

Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Journal of Educational Technology & Society, 15*(3), 42–57. Geraadpleegd van https://www.jstor.org/stable/pdf/jeductechsoci.15.3.42.pdf

Gurung, R. A. R., Weidert, J., & Jeske, A. (2010). Focusing on How Students Study. *Journal of the Scholarship of Teaching and Learning, 10*(1), 28-35. Geraadpleegd van https://scholarworks.iu.edu/journals/index.php/josotl/article/view/1734

Hogeschool van Amsterdam (2018). *Jaarverslag 2018*. Geraadpleegd op 7 november 2019, van https://www.amsterdamuas.com/binaries/content/assets/hva/over-de-hva/feiten-en-cijfers/jaarverslagen-vanaf-2016/hva_jaarverslag_2018.pdf?1561457394197

Kirby, A., & McElroy, B. (2003). The effect of attendance on grade for first year economics students in University College Cork. *The Economic and Social Review, Economic and Social Studies, 34*(3), 311-326. Geraadpleegd van https://www.esr.ie/ESR_papers/vol34_3/Vol34_3Kirby.pdf

Larrabee Sønderlund, A., Hughes, E., & Smith, J. (2019). The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology, 50*(5), 2594-2618. https://doi.org/10.1111/bjet.12720

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *The American Behavioral Scientist (Beverly Hills), 57*(10), 1439-1459. doi:10.1177/0002764213479367

Lowes, S., Lin, P., & Kinghorn, B. (2015). Exploring the link between online behaviours and course performance in asynchronous online high school courses. *Journal of Learning Analytics*, *2*(2), 169-194. Geraadpleegd van http://dx.doi.org/10.18608/jla.2015.22.13

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers and Education, 54*(2), 588-599. doi:10.1016/j.compedu.2009.09.008

McCoy, S., & Byrne, D. (2017). Student retention in higher education. In J. Cullinan & D. Flannery (Eds), *Economic Insights on Higher Education Policy in Ireland: Evidence from a Public System* (pp. 111–141). doi:10.1007/978-3-319-48553-9_5

Milne, J., Jeffrey, L. M., Suddaby, G., & Higgins, A. (2012). Early identification of students at risk of failing. In *Australian Society for Computers in Learning in Tertiary Education Annual Conference (ASCILITE)* (Vol. 1, pp. 25-28). Geraadpleegd van https://www.ascilite.org/conferences/Wellington12/2012/images/custom/milne,_john_-_early_identification.pdf

Morris, L. V., Finnegan, C., & Wu, S. (2005). Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education, 8*(3), 221-231. doi:10.1016/j.iheduc.2005.06.009

Norris, D. M., & Baer, L. L. (2013). *Building organizational capacity for analytics. EDUCAUSE.* Geraadpleegd van https://library.educause.edu/resources/2013/2/building-organizational-capacity-for-analytics

Paige, S. M., Wall, A. A., Marren, J. J., Dubenion, B., & Rockwell, A. (2017). *The learning community experience in higher education: High-impact practice for student retention.* New York: Routledge.

Rafaeli, S., & Ravid, G. (1997). Online, web-based learning environment for an information systems

course: Access logs, linearity and performance. *ISECON, 97,* 92-99. Geraadpleegd van http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=DF5E9EFD69C10F3C5813E29ABEC2291E?doi=10.1.1.16.9119&rep=rep1&type=pdf

Rodgers, J. R. (2001). A panel-data study of the effect of student attendance on university performance. *Australian Journal of Education*, *45*(3), 284-295. doi:10.1177/000494410104500306

Ryabov, I. (2012). The effect of time online on grades in online sociology courses. *MERLOT Journal of Online Learning and Teaching*, *8*(1), 13-23. Geraadpleegd van https://jolt.merlot.org/vol8no1/ryabov_0312.pdf

Scheffel, M., Drachsler, H., Stoyanov, S., & Specht, M. (2014). Quality indicators for learning analytics. *Journal of Educational Technology & Society*, *17*(4), 117-132. Geraadpleegd van https://pdfs.semanticscholar.org/e433/4e54049a37beca208bd7aa4b4b57cc5f551d.pdf?_ga=2.176363365.1391181403.1604521170-670125675.1604521170

Sieben, I., & Linssen, L. (2009). *Logistische regressie analyse: een handleiding.* Geraadpleegd van https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwichsmx5uTvAhWS6aQKHSSQAJIQFjAAegQIAxAD&url=https%3A%2F%2Fwww.ru.nl%2Fpublish%2Fpages%2F771745%2Flogistischeregressie.pdf&usg=AOvVaw3UrphgMzLkAcI21jEF6Y2C

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380-1400. doi:10.1177/0002764213498851

Siemens, G., & Baker, R. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252-254). doi:10.1145/2330601.2330661

SURF (2019). *Begrippenkader studiedata en learning analytics*. Geraadpleegd van https://www.surf.nl/kennisdossier-benutten-van-studiedata/begrippenkader-studiedata-en-learning-analytics

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, *45*(1), 89-125. Geraadpleegd van https://www.jstor.org/stable/1170024?seq=1#metadata_info_tab_contents

Van Berkel, H., Jansen, E., & Bax, A. (2012). *Studiesucces bevorderen: het kan en is niet moeilijk.* Amsterdam: Boom Digitale Uitgevers.

Van den Bogaard, M. E. D., & De Vries, P. (2017). "Learning Analytics is about Learning, not about Analytics" A reflection on the current state of affairs. In *45th Annual SEFI Conference* ISEP Lisbon. Geraadpleegd van https://research.tudelft.nl/en/publications/learning-analytics-is-about-learning-not-about-analytics-a-reflec

Van den Ende, I., De Vreede, R., Zandvliet, K., & De Vleeschouwer, E. (2019) *Het bindend*

studieadvies in het hoger onderwijs. Rapportage in opdracht van het ministerie van OCW. Geraadpleegd van https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/brochures/2019/04/04/het-bindend-studieadvies-in-het-hoger-onderwijs/het-bindend-studieadvies-in-het-hoger-onderwijs.pdf.

Vereniging Hogescholen (2019). *Professionals voor morgen. Strategische agenda Vereniging Hogescholen 2019 – 2023.* Geraadpleegd van https://www.verenigingh ogescholen.nl/kennisbank/professionals-voor-morgen-strategische-agenda-vereniging-hogescholen-2019-2023/

You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *The Internet and Higher Education, 29*, 23-30. https://doi.org/10.1016/j.iheduc.2015.11.003

Yu, T., & Jo, I. H. (2014). Educational technology approach toward learning analytics: Relationship between student online behavior and learning performance in higher education. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (pp. 269-270). doi:10.1145/2567574.2567594

Bijlagen

Bijlage 1. Modellen voor hiërarchische logistische regressieanalyse (over de gehele onderwijsperiode – totaal variabelen)

Tabel

Model 1: Omnibus test van de predictor aanwezigheid

	Chi-square	df	Sig.
Step	27.49	1	.000
Block	27.49	1	.000
Model	27.49	1	.000

Tabel

Model 2: Omnibus test van de predictoren aanwezigheid en aantal bekeken pagina's

	Chi-square	df	Sig.
Step	6.54	1	.011
Block	6.54	1	.011
Model	34.02	2	.000

Tabel

Model 3: Omnibus test van de predictoren aanwezigheid, aantal bekeken pagina's en totale tijdsduur online

	Chi-square	df	Sig.
Step	.10	1	.752
Block	.10	1	.752
Model	34.12	3	.000

Bijlage 2. Output logistische regressie analyse (over de gehele onderwijsperiode – totaal variabelen)

Tabel

Model summary

-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
416.454[a]	.093	.128

Tabel

Classification tabel: geen predictoren, alleen de constante

		Predicted		
		Module behaald		
Observed		niet behaald	behaald	Percentage Correct
Module behaald	niet behaald	0	121	.0
	behaald	0	228	100.0
Overall Percentage				65.3

Tabel

Classification tabel: voorspellende waarde met aanwezigheid en aantal bekeken pagina's

		Predicted		
		Module behaald		Percentage
Observed		niet behaald	behaald	Correct
Vak behaald	niet behaald	33	88	27.3
	behaald	18	210	92.1
Overall Percentage				69.6

Tabel

Variables in the equation

							95% C.I.for EXP(B)	
	B	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
Aanwezigheid	2.393	.693	11.920	1	.001	10.943	2.813	42.561
Pagina_totaalbezocht	.059	.023	6.494	1	.011	1.061	1.014	1.110
Constant	-1.761	.455	15.004	1	.000	.172		

Tabel

Controleren op lineariteit: Hosmer and Lemeshow Test

Chi-square	df	Sig.
3.999	8	.857

Tabel

Controleren op multicollineairteit

	Collinearity Statistics	
	Tolerance	VIF
Aanwezigheid	.781	1.281
Pages_bezochttotaal	.781	1.281

a. Dependent Variable: Module behaald

Bijlage 3. Modellen voor hiërarchische logistische regressieanalyse (over de eerste drie lesweken)

Tabel

Model 1: Omnibus test van de predictor aanwezigheid lesweek 1 t/m 3

	Chi-square	df	Sig.
Step	22.381	1	.000
Block	22.381	1	.000
Model	22.381	1	.000

Tabel

Model 2: Omnibus test van de predictoren aanwezigheid en aantal bekeken pagina's lesweek 1 t/m 3

	Chi-square	df	Sig.
Step	1.155	1	.283
Block	1.155	1	.283
Model	23.536	2	.000

Tabel

Model 3: Omnibus test van de predictoren aanwezigheid, aantal bekeken pagina's en totale tijdsduur online lesweek 1 t/m 3

	Chi-square	df	Sig.
Step	1.169	1	.280
Block	1.169	1	.280
Model	24.705	3	.000

Bijlage 4. Output logistische regressie analyse (over de eerste drie lesweken)

Tabel

Model summary

-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
428.096[a]	.062	.086

Tabel

Classification table: voorspellende waarde met predictor aanwezigheid

		Predicted		
		Module behaald		Percentage
Observed		niet behaald	behaald	Correct
Vak behaald	niet behaald	33	88	27.3
	behaald	21	207	90.8
Overall Percentage				68.8

Tabel

Variables in the equation

							95% C.I.for EXP(B)	
	B	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
Proportie aanwezig week 1 tm 3	3.015	.668	20.363	1	.000	20.395	5.505	75.563
Constant	-2.026	.600	11.404	1	.001	.132		